

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/126122/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Moat, Stuart J., Schulenburg-Brand, Danja, Lemonde, Hugh, Bonham, James R., Weykamp, Cas W., Mei, Joanne V., Shortland, Graham S. and Carling, Rachel S. 2020. Performance of laboratory tests used to measure blood phenylalanine for the monitoring of patients with phenylketonuria. *Journal of Inherited Metabolic Disease* 43 (2) , pp. 179-188.10.1002/jimd.12163 file

Publishers page: <http://dx.doi.org/10.1002/jimd.12163>
<<http://dx.doi.org/10.1002/jimd.12163>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Performance of laboratory tests used to measure blood phenylalanine for the monitoring of patients with Phenylketonuria

Moat SJ^{1,2*}, Schulenburg-Brand D¹, Lemonde H³, Bonham JR⁴, Weykamp CW⁵,

Mei JV⁶, Shortland GS⁷, Carling RS^{8,9}

¹Department of Medical Biochemistry, Immunology & Toxicology, ²School of Medicine, Cardiff University, University Hospital Wales, Cardiff, CF14 4XW, UK. ³Paediatric Metabolic Medicine, Evelina Children's Hospital, Guys & St Thomas' NHSFT, London, UK. ⁴Department of Clinical Chemistry, Sheffield Children's (NHS) FT, Sheffield, UK. ⁵MCA Laboratory, Queen Beatrix Hospital, Winterswijk, The Netherlands. ⁶Newborn Screening and Molecular Biology Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA. ⁷Department of Child Health, University Hospital Wales, Cardiff, UK. ⁸Biochemical Sciences, Viapath, Guys & St Thomas' NHSFT, ⁹GKT School of Medical Education, King's College, London.

*Corresponding author: Prof Stuart J Moat, Department of Medical Biochemistry, Immunology & Toxicology, University Hospital Wales, Cardiff, CF14 4XW. Stuart.Moat@Wales.nhs.uk, Tel +44 29 20 743 562, Fax +44 29 20 744 065.

Abstract - 245 words

Total word count = 4658

Figures = 3, Tables = 3

Funding – NA

Ethical approval – NA

Key words; Phenylketonuria, monitoring, treatment-ranges, dried blood spots, precision, accuracy, bias

Abbreviation List;

Analytical coefficient of variation (CV_A), within person biological variation (CV_I), Clinical & Laboratory Standards Institute (CLSI), Centers for Disease Control and Prevention (CDC), certified reference material (CRM), Dried blood spot (DBS), European (ERNDIM), External Quality Assessment (EQA), flow-injection analysis tandem mass-spectrometry (FIA-MS/MS), High

Performance Liquid Chromatography (HPLC), ion exchange chromatography (IEC), measurement of uncertainty (MU) , National Institute of Standards & Technology (NIST), phenylketonuria (PKU), reference change value (RCV), standard deviation (SD), total allowable error (TAE), total error (TE).

Compliance with Ethics Guidelines

Stuart Moat declares that he has no conflict of interest.

Danja Schulenburg-Brand declares that she has no conflict of interest.

Hugh Lemonde declares that he has no conflict of interest.

James Bonham declares that he has no conflict of interest.

Cas Weykamp declares that he has no conflict of interest.

Joanne Mei declares that she has no conflict of interest.

Graham Shortland declares that he has no conflict of interest.

Rachel Carling declares that she has no conflict of interest.

This article does not contain any studies with human or animal subjects performed by the any of the authors.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service, or the U.S. Department of Health and Human Services.

Contributions of individual authors -

SJM was the lead author of the manuscript. All authors made contributions to the perspective/review, and/or acquisition of data, and/or analysis and interpretation of data.

All authors participated in drafting the article and revising it critically for important intellectual content. All authors gave approval of the final version submitted.

Abstract

Analysis of blood phenylalanine is central to the monitoring of patients with phenylketonuria (PKU) and age-related phenylalanine target treatment-ranges (0-12 years; 120–360 μ mol/L, and >12 years; 120-600 μ mol/L) are recommended in order to prevent adverse neurological outcomes. These target treatment-ranges are based upon plasma phenylalanine concentrations. However, patients are routinely monitored using dried bloodspot (DBS) specimens due to the convenience of collection. Significant differences exist between phenylalanine concentrations in plasma and DBS, with phenylalanine concentrations in DBS specimens analysed by flow-injection analysis tandem mass spectrometry (FIA-MS/MS) reported to be 18-28% lower than paired plasma concentrations analysed using ion-exchange chromatography (IEC). DBS specimens with phenylalanine concentrations of 360 μ mol/L and 600 μ mol/L, at the critical upper-target treatment-range thresholds would be plasma equivalents of 461 μ mol/L and 768 μ mol/L respectively, when a reported difference of 28% is taken into account. Furthermore, analytical test imprecision and bias in conjunction with pre-analytical factors such as volume and quality of blood applied to filter paper collection devices to produce DBS specimens affect the final test results. Reporting of inaccurate patient results when comparing DBS results to target treatment-ranges based on plasma concentrations, together with inter-laboratory imprecision could have a significant impact on patient management resulting in inappropriate dietary change and potentially adverse patient outcomes. This review is intended to provide perspective on the issues related to the measurement of phenylalanine in blood specimens and to provide direction for the future needs of PKU patients to ensure reliable monitoring of metabolic control using the target treatment-ranges.

Synopsis - Utility of the target treatment-ranges for monitoring patients with Phenylketonuria (PKU) is limited by pre-analytical factors and the analytical performance of existing laboratory assays for blood phenylalanine.

Introduction

Patients with Phenylketonuria (PKU) (OMIM #261600) are managed by the use of a phenylalanine restricted diet (in conjunction with protein substitute supplements) to lower the blood phenylalanine concentrations and prevent the adverse neurological sequelae (Scriver et al 2001). Several guidelines for the diagnosis and management of patients with PKU have been published over the years (Smith 1993; Wappner et al 1999; Vockley et al 2014; Singh et al 2014; van Spronsen et al 2017). One of the key recommendations is the monitoring of the phenylalanine restricted diet, using appropriate age-related phenylalanine target treatment-ranges to prevent the adverse neurological outcomes. In the latest set of guidelines (van Spronsen et al 2017), the following target treatment-ranges are recommended; 120-360 μ mol/L for individuals aged 0-12 years and 120-600 μ mol/L for individuals older than 12 years. It should be noted that the grade of recommendation for these target ranges is graded as category D (case series reports & expert opinions). Optimal target ranges have also been recommended for women trying to conceive and during pregnancy (120-360 μ mol/L). An infant born to a PKU mother who is not achieving optimal control can result in neurological deficits, microcephaly and congenital heart defects. Accurate and reproducible monitoring of blood phenylalanine concentrations is therefore critical to the management of PKU. This review is intended to provide perspective on the issues related to the measurement of phenylalanine in blood specimens and to provide direction for the future needs of PKU patients in terms of accurate monitoring of metabolic control. With the evidence presented in this review, both laboratory scientists and clinicians should consider whether the analytical performance of the methods routinely used in their laboratory is adequate to support the use of the recommended patient target treatment-ranges in the management of patients with PKU.

Laboratory methods used to measure blood phenylalanine

Traditionally, the standard method for the laboratory diagnosis and monitoring of PKU patients is by quantifying phenylalanine in deproteinised plasma samples, using ion exchange chromatography (IEC) with ninhydrin detection. Whilst this method is reproducible, it has a long analytical run time (~2 hours per sample), low sample throughput and lacks specificity in comparison with liquid chromatography tandem mass spectrometry (LC-MS/MS). These limitations are reflected in the number of laboratories moving away from IEC to LC-MS/MS in recent years (Carling 2018a).

Irrespective of technique, the measurement of plasma amino acids can be performed with reasonable precision, with a typical intra-laboratory coefficient of variation (CV) for phenylalanine being approximately 5%. The CV is a measure of the variability of the test results and is the ratio of the standard deviation (SD) to the mean of the results obtained multiplied by 100 (i.e. $CV (\%) = SD/mean \times 100$). As expected, the inter-laboratory variation is greater, as shown by data from the European Research Network in Inherited Metabolic Diseases (ERNDIM) Quantitative Plasma Amino Acids External Quality Assessment (EQA) scheme, with an average CV of 9.5%, at a nominal concentration of 355 $\mu\text{mol/L}$ ($n=282$ participants), and this can be attributed predominantly to the variation in the test standardisation, compounded by differences in laboratory practice (Carling 2018b). Generally, an aqueous calibrator is used to standardise the test and these tend to be produced in-house, although an increasing number of laboratories now use a certified reference material (CRM), to standardise assays; the Sigma TraceCert solution and the National Institute of Standards & Technology (NIST) SRM2389a solution, both of which are available commercially. It should be noted that these certified reference materials (CRM) are aqueous solutions of amino acids and not matrix (plasma) matched.

Plasma amino acid analysis is impractical for routine monitoring of metabolic control due to the logistics required in collecting a venous blood specimen twice weekly to monthly (depending on age and clinical need) from every patient, including very young children. Plasma amino acid analysis undertaken in this way has been used primarily for diagnosis and clinical situations, in which a full profile of amino acids is informative, e.g. complete nutritional assessment. Instead, measurement of phenylalanine in dried bloodspot (DBS) specimens is widely favoured due to the convenience of collecting blood from a finger-prick onto filter paper in the patient's home and mailing the sample directly to the laboratory. Monitoring patients using DBS has been routinely done by a variety of methods since the inception of screening for PKU. Methods used include: bacterial inhibition assay, fluorimetry, IEC, high performance liquid chromatography (HPLC) and enzymatic analysis (Belton et al 1973; Rudy et al 1987; Rohr et al 1996; Wendel et al 1991). During the last 25 years flow-injection analysis tandem mass-spectrometry (FIA-MS/MS) has been used to analyse DBS for patient monitoring, following the introduction of this technology for routine newborn screening of PKU (Chace et al 1993). Measurement by FIA-MS/MS has a shorter analysis time and high sample throughput; results can therefore be communicated to patients in a timely manner allowing prompt dietary adjustment.

FIA-MS/MS is inherently limited by lack of specificity as analytes are not separated chromatographically. Instead, specificity is achieved by multiple reaction monitoring, which allows rapid and continuous monitoring of the specific daughter ions relating to the analyte of interest, hence any isobaric compound with a common daughter ion has the potential to interfere e.g. benzocaine (anaesthetic agent found in some antiseptic wipes which may be used to clean skin prior to sample collection) (Williams et al 2015), lorazepam and omeprazole (<https://massbank.eu/MassBank/>). When these assays were originally established, FIA-MS/MS

proved to be a robust assay with rapid throughput (1.5 minutes per sample) and adequate sensitivity and specificity. We established a snapshot of current FIA-MS/MS analytical performance by circulating a commercially available DBS QC material to 16 UK laboratories. The mean value obtained was 174 μ mol/L (range 100 to 256 μ mol/L) with an inter-laboratory CV of 20.2%. The assigned target-value of the material (167 μ mol/L) also highlighted the large biases displayed by individual laboratories, which ranged from -40% to +53%. Technology has evolved in recent years and modern mass-spectrometers now have the ability to scan faster. These changes, in conjunction with the introduction of ultra performance liquid chromatography, mean that laboratories now have the capability to routinely analyse phenylalanine in DBS with superior specificity and precision, and with a comparable analysis time and robustness.

A further limitation to the utility of DBS specimens for monitoring patients is the lack of a commercially available CRM for DBS amino acids on which to standardise laboratory tests. As a result, DBS calibrators tend to be produced in-house by collecting blood from a healthy donor and adding an aqueous phenylalanine enrichment prior to application onto filter paper. The exact preparation of the DBS calibrator varies between laboratories e.g. volume of blood added to the filter paper, varying haematocrit of the specimen, or use of lysed blood specimens can all affect the measured concentration (George & Moat 2016; Lawson et al 2016; Mei et al 2001; Hall et al 2015). More importantly, the method used to assign DBS calibrator values can influence the analytical result.

Derivation of the target treatment-ranges for ‘blood’ phenylalanine and the differences between DBS and plasma phenylalanine results

It should be recognised that the vast proportion of studies evaluating the neurological outcomes in patients with PKU, are mostly based upon plasma phenylalanine concentrations (Vockley et al 2014; Stroup et al 2016; Van Spronsen et al 2017) and these studies form the basis of the recommended target treatment-ranges. However, these guidelines refer to the monitoring of ‘blood’ phenylalanine concentrations, but do not make any distinction between the two specimen types (Smith 1993; Wappner et al 1999; Vockley et al 2014; Singh et al 2014), or refer to the observed differences, but then do not provide guidance as to the most appropriate specimen type to be used (van Spronsen et al 2017). This is disconcerting in that there are significant differences between plasma phenylalanine measured by IEC and DBS phenylalanine measured by FIA-MS/MS, with concentrations in DBS being up to 18-28% lower (Table 1). The difference between DBS and plasma phenylalanine concentrations is reduced from -28% to -15% when both plasma and DBS are analysed using IEC (Stroup et al 2016). Furthermore, the difference between liquid whole blood and plasma phenylalanine concentrations is ~8% when measured by the same laboratory technology (Table 1). These findings suggest that the reported differences between plasma and DBS specimens are due to several factors; distribution of phenylalanine between plasma and erythrocytes, extraction efficiency of phenylalanine from DBS and laboratory instrument test biases. Using DBS collected by trained healthcare professionals, we observed that DBS phenylalanine concentrations measured with FIA-MS/MS are 18% lower relative to the paired plasma concentrations measured by IEC (Figure 1), which is similar to those reported previously. The negative bias of 18% in our laboratory is reproducible and stable over a long period of time.

The differences between DBS and plasma phenylalanine concentrations could have a significant clinical impact in terms of management. For example; it is recommended that during pregnancy 'blood' phenylalanine concentrations should be maintained between 120 to 360 μ mol/L. Assuming these target ranges are based upon plasma concentrations, but the patient is monitored using DBS concentrations, a DBS concentration of 360 μ mol/L is equivalent to a plasma concentration of 425 μ mol/L in our laboratory based upon current performance. Whereas, in a laboratory with a negative bias of 28%, DBS specimens with phenylalanine concentrations of 360 μ mol/L and 600 μ mol/L, at the upper critical target treatment-range thresholds would be plasma equivalents of 461 μ mol/L and 768 μ mol/L respectively (Figure 2).

Additional factors that affect patient phenylalanine test results

Although the published patient treatment and monitoring guidelines emphasise the importance of biochemical monitoring of phenylalanine concentrations, the impact of the following factors on the final result reported are not considered in any of the guidelines:

- The significance of the analytical test variation (imprecision) and analytical test bias (inaccuracy) in the measurement of phenylalanine.
- The total allowable error for a phenylalanine test that can be tolerated before it has a significant impact upon clinical decision making.
- The impact of the pre-analytical factors such as the amount or volume of blood applied to the filter paper to produce DBS specimens on the phenylalanine results obtained.

Analytical test variation (Imprecision)

Laboratory tests are not perfect and it is imperative that healthcare professionals understand the factors that affect the phenylalanine test results that they receive and act on. A requirement for the accreditation of clinical laboratories according to the International Standards Organisation (ISO) 15189, is to provide the measurement of uncertainty (MU) for each test performed (ISO 15189:2012). MU is defined as the value that is associated with the test result that describes the dispersion of values that could be obtained for the test result due to the uncertainties arising within the test procedure. MU is calculated using quality control (QC) specimens analysed in every batch of patient samples over a period of time to encompass the combined effect of all the analytical factors that influence the test result (Wayne 2012). In one author's laboratory (SJM) the mean and standard deviation (SD) of a DBS QC specimen analysed in 166 separate test batches over an 8

month period using FIA-MS/MS was 361 (19.3) $\mu\text{mol/L}$, corresponding to an analytical inter-assay CV of 5.3%. Using these data, the MU of the test can be calculated using the following equation:

$$\text{MU} = \text{SD} \times 1.96.$$

Using a coverage factor of 1.96, there is a 95% chance that the true result lies within a range covered by the result value \pm MU. The calculated MU for the DBS phenylalanine test is 37.8 $\mu\text{mol/L}$ or 10.5%. This means that for a DBS specimen with a phenylalanine concentration at the critical upper treatment threshold of 360 $\mu\text{mol/L}$, we are 95% confident that the result is between 322.2 and 397.8 $\mu\text{mol/L}$ (i.e. 360 $\mu\text{mol/L} \pm 10.5\%$). However, this concentration range does not include variability due to the pre-analytical factors such as difference in DBS specimen collection techniques or the inherent differences between different laboratory test methodologies. The inter-assay CV and MU for the plasma assay in our laboratory is 3% and 5.8% (i.e. 1.96×3) respectively.

Understanding the variability of the test used to monitor patients serially in order to assess optimal dietary control is also important and the reference change value (RCV) can be used to determine this. The RCV refers to the minimum critical difference between two consecutive results, in the same patient, which needs to be exceeded in order for a significant change to have occurred (Fraser 2009). As there are two sets of phenylalanine results to consider, we must take into consideration two sets of variations that must be combined to produce the RCV. This combined variation is represented by the following equation:

$$\text{RCV} = 2.77 \times \sqrt{\text{CV}_A^2 + \text{CV}_I^2}$$

Where CV_A = the analytical coefficient of variation (CV) of the test and CV_I = within person biological variation. The CV_I of phenylalanine in healthy adult individuals is 9.5% (Corte & Venta

2010) and the CV_A (SJM – laboratory) for a DBS phenylalanine QC is 5.3%. The two sets of results being compared need to be greater than 2.77 (i.e. $\sqrt{2} \times 1.96$) times the analytical and within-person biological variations. The calculated RCV for phenylalanine using DBS specimens is 30.2%. Therefore, for a patient with a DBS phenylalanine concentration of $360\mu\text{mol/L}$, an increase or decrease of $>109\mu\text{mol/L}$, would have to occur before the results are considered significantly different (95% confidence) i.e. not due to the within-person biological variation or test variation.

It could be argued that including the effect of biological variation within the RCV calculation is not applicable as PKU patients are prescribed a specialist diet and supplements in a manner to maintain stable phenylalanine concentrations throughout the day (MacDonald et al 1996; van Spronsen et al 2017). Furthermore, patients are advised to collect specimens at the same time of day, usually fasting in the morning, when the phenylalanine concentration is highest (MacDonald et al 1996). Therefore, assuming optimal compliance with this regimen, the use of the RCV calculation may not be an appropriate approach to assess serial changes in phenylalanine concentrations in a patient, as only large variations may be deemed as a significant change. An alternative approach to calculate the critical difference between two patient results would be to remove the contribution of the within-person biological variation. This can be calculated by multiplying the assay imprecision (5.3%) by 2.77 which equals 14.7%. Therefore at a concentration of $360\mu\text{mol/L}$, an increase or decrease of $>52.9\mu\text{mol/L}$ in consecutive DBS specimens would be considered significant. However, it should be noted that the MU and RCV calculations do not take into account the test bias. The RCV value using plasma test results is 27.6% and 8.3% (i.e. $2.77 \times 3\%$) when the contribution of biological variation is removed.

Analytical test bias (Inaccuracy)

Test bias refers to the difference between the test result obtained and the reference or 'true' value. Clinical laboratories compare their analytical performance with results from EQA schemes (Table 2). In an attempt to define what the acceptable test bias of phenylalanine should be, the biological variation of phenylalanine can be used to derive the acceptable bias, and this has been calculated to be 10.4% (Corte & Venta 2010). Furthermore, the Horwitz equation can also be used to predict inter-laboratory variation on the basis of analyte concentration alone, as it is independent of method, specimen matrix and analyte. The Horwitz equation (%) = $2C^{-0.15}$, where C is the concentration of the analyte, expressed as a mass fraction (Horwitz & Albert 2006). For phenylalanine in the concentration range of 10 – 500µmol/L, this would equate to a target inter-laboratory variation of approximately 10%. It is clear that the plasma tests conform to this calculated acceptable bias of 10%, whereas the DBS tests are highly variable depending on the EQA scheme (Table 2) and this is in part due to the fact that there is no CRM commercially available for DBS phenylalanine tests. Interestingly, the target test bias of 10.4% derived from biological variation is similar to that of 10% calculated using the Horwitz formulae:

The UK National External Quality Assessment Service (NEQAS) and the Centers for Disease Control and Prevention (CDC) schemes have been available for many years and are intended to assess the performance of the newborn screening programmes. More recently ERNDIM has introduced a DBS scheme targeted towards an assessment of control in patients receiving treatment. The pilot scheme was operational in 2017 and 2018 and supplied 79 laboratories with four DBS specimens (phenylalanine range 120 to 940µmol/L) and a DBS phenylalanine calibrator (350µmol/L). Laboratories were asked to submit results from both the DBS specimens and the DBS calibrator. Results from the four specimens were then virtually "calibrated" for each laboratory using the

standard material. The inter-laboratory CVs from the “raw” and the “calibrated” results were calculated for the four specimens (Table 3). It can be seen that the mean inter-laboratory CV is improved by the use of this virtual standard from 20% using the “raw” data to 10% when the “calibrated” data is analysed. This effect is evident at all concentrations from 120 to 940 μ mol/L (Bonham & Weykamp 2018). Effectively, inter-laboratory variation could be reduced significantly by introducing a common DBS calibrator. In order to address the issues regarding the inter-laboratory biases observed between newborn screening laboratories, a European working calibrator for DBS phenylalanine (EWS-Phe-01) was created (Dhondt et al 1998) and this is now produced by the International Society for Newborn Screening. A multi-analyte DBS amino acid reference material has also been produced by the CDC in the US (Chace et al 1999). However, these materials are only available in limited quantities to kit manufacturers and EQA scheme organisers.

Reporting inaccurate monitoring results could have profound effects in that patients may be falsely reassured with lower results, where laboratories have a negative bias for DBS phenylalanine and conversely, those laboratories with a positive bias providing falsely elevated results which may prompt a stricter dietary regimen, which may lead to non-compliance issues. Therefore, with such large and variable biases for DBS phenylalanine results being observed between different laboratories, it is evident that consideration should be given to test bias when utilising the treatment-ranges.

Establishing optimal analytical performance targets for phenylalanine tests

Unfortunately, there is not a single simple approach to establishing optimal analytical targets for analytes measured in the clinical laboratory and several strategies exist (Kenny et al 1999). The total allowable error (TAE) of a test can be used to establish performance targets for a test. The TAE of a test is the maximum error that can be tolerated for a test before it has a significant impact upon clinical decision making. TAE is not based on the performance of the analytical test being used, but is dependent on the clinical use of the test result and the inherent biological variability of the analyte (Klee 2010). TAE should be derived objectively from an analysis of studies assessing the clinical impact of the test performance, although this is often difficult to achieve.

Other strategies include professional recommendations by expert groups or those derived from data obtained from biological variation. The TAE of 18.2% and a bias of 10.4% for phenylalanine have been calculated using data from the biological variation of amino acids in plasma from healthy adult subjects (Corte & Venta 2010). The total error (TE) of the phenylalanine tests can be calculated using the following formulae:

$$TE = \text{Test bias} + (1.96 \times \text{test CV})$$

Table 2 shows the TE of the plasma and DBS tests in our laboratories. On comparing these to the TAE of 18.2%, it would suggest that the plasma phenylalanine tests are clinically fit for purpose to monitor patients. Conversely, the performance of the DBS tests is less consistent, depending on which EQA results are used to calculate the test bias (Table 2).

To date no study has derived an analytical goal from studies that have assessed the clinical impact of test performance on clinical outcomes in patients with PKU. However, it is possible to derive a

TAE using clinical outcome studies. Data from a meta-analysis of 17 studies (n=432 patients), demonstrated a threshold effect of a phenylalanine concentration $>400\mu\text{mol/L}$, that was associated with an IQ of less than 85 in patients aged <6 years (Fonnesbeck et al 2013). Furthermore, it was also shown that those patients (age range 8-13 years) with a lifetime phenylalanine $>400\mu\text{mol/L}$, did worse than those with a phenylalanine $<400\mu\text{mol/L}$ in all tests assessing executive function (Leuzzi et al 2004). It is therefore imperative that laboratory tests must be able to distinguish with confidence between a test result of $360\mu\text{mol/L}$ (upper-limit of the target treatment-range) and one of $400\mu\text{mol/L}$, which is potentially damaging. Therefore the measurement error of the test must not exceed the difference between the two test results (i.e. $40\mu\text{mol/L}$), which relates to a TAE for the test of 11.1% (i.e. $(40/360) \times 100$). This is significantly lower than the TAE of 18.1% derived by biological variation. However, it is recognised that using biological variation to derive TAE can lead to over estimation (Oosterhuis et al 2011). If we look at whether the test is able to confidently detect a change between 360 and $300\mu\text{mol/L}$, the TAE would be 16.7% (i.e. $(60/360) \times 100$), which is closer to the TAE of 18.2% derived by biological variation. Therefore, when comparing patient phenylalanine test results to target treatment-ranges we must understand the TE of the test and whether or not it is fit for purpose i.e. the TE of the test used should ideally be less than the TAE of the test.

Pre-analytical factors affecting the measurement of phenylalanine in DBS

The size (i.e. volume of blood applied to the filter paper) and quality of the bloodspot has a significant impact on the results obtained. Phenylalanine concentrations are significantly lower in smaller DBS relative to larger DBS specimens (Hall et al 2015; George & Moat 2016; Lawson et al 2016). If a liquid blood specimen, containing phenylalanine at a concentration of $360\mu\text{mol/L}$, is applied to filter paper in different volumes (10 to $100\mu\text{L}$) the mean concentration of the actual measured phenylalanine in the DBS samples (using a standard 3.2mm sub-punch for analysis) would vary from $300\mu\text{mol/L}$ in the $10\mu\text{L}$ DBS to $400\mu\text{mol/L}$ in the $100\mu\text{L}$ DBS (Figure 3). If the effect of DBS size and the MU of the test are combined, then the range of results that could be reasonably observed for a specimen with a concentration of $360\mu\text{mol/L}$ is anything between 269 and $442\mu\text{mol/L}$. However, this does not take into consideration the effect of the variable biases observed for the phenylalanine results, which may further compound these pre-analytical errors. A comparable effect is also seen for tyrosine concentrations (George & Moat 2016), which can be simultaneously measured in the FIA-MS/MS DBS phenylalanine assay, and is used to monitor the nutritional supplementation of tyrosine.

Following the reported evidence that small and poor quality DBS specimens produce falsely low and erroneous results for the metabolites used for newborn screening (George & Moat 2016; Lawson et al 2016), UK newborn screening guidelines for DBS specimen rejection/acceptance were developed. Several UK metabolic laboratories also adopted these specimen acceptance/rejection criteria for DBS specimens received for monitoring of PKU patients. Prior to implementation, an audit was undertaken to assess the impact on the DBS monitoring service. It was demonstrated in one of the author's laboratory (RSC) that if these criteria were applied to specimens received for PKU monitoring over a 4 week period, 143 out of the 300 specimens (47.7%) received would have been

rejected. The vast majority of these specimens were too small and would have produced significant negative biases. This finding highlights the fact that the negative bias observed due to small DBS and the lower concentration in DBS versus plasma specimens, would result in falsely lower results, thereby giving false re-assurances as to optimal metabolic control in patients. Following patient / parent training for the collection of DBS specimens the number rejected due to being too small or poor quality reduced to 26.2%.

Filter paper collection devices for capillary blood collection from heel or finger pricks are Class II Medical Devices (21 CFR 862.1675), and should meet international criteria for performance by the Clinical & Laboratory Standards Institute (CLSI 2013). The Newborn Screening Quality Assurance Program at CDC conducts voluntary evaluations of all lots of Food and Drug Administration (FDA)-registered collection devices before they are released to the user community for newborn screening and other applications (Mei et al 2001; Mei et al 2010). In general, all filter paper lots comply with CLSI performance criteria (Mei et al 2010). The CLSI provides the framework for filter paper performance. Because filter paper is a natural product produced from cotton linters, defining how the matrix influences blood collection is important so that precision and reproducibility can be achieved from lot-to-lot (Mei 2010). These international standardization efforts ensure uniformity of specimen collection, calibrators, QC and reference materials for newborn screening assays. Using DBS specimens for patient monitoring adds additional requirements for the precision and accuracy of analyte recovery. The type of matrix used for calibration and QC materials will influence the analyte recovery. Ideally, methods testing patient DBS specimens should also use DBS calibration and QC to correct for the filter paper matrix.

Discussion and Recommendations

Although the difference between DBS and plasma phenylalanine concentrations in paired specimens has previously been reported, the clinical impact on patient management has not been assessed. A DBS result compared to a plasma target treatment-range could be falsely reassuring and potentially damaging, particularly in those laboratories where large negative biases are observed for DBS concentrations. To provide comparable results for patient monitoring, a calibration factor could be used to report DBS results as plasma equivalents to ensure meaningful comparison of results to the recommended target treatment-ranges. This is preferable to reporting patient results against different target treatment-ranges as this may cause confusion for both patients and clinicians. However, it is essential that the patient DBS specimens collected are of adequate size and quality to ensure accurate results, because the differences between DBS and plasma specimens are more variable when patients apply their own blood onto the filter paper collection devices compared to those applied by trained healthcare professionals (Stroup et al 2016). Improvement in DBS specimen quality could be potentially achieved by the use of blood collection devices, that collect defined volumes of liquid blood for sampling (Lenk et al 2015; Leuthold et al 2015; Neto et al 2018; Spooner et al 2015) and such devices should be evaluated in order to improve the biochemical monitoring of patients. It has been highlighted that the development of point-of-care devices to measure liquid whole blood phenylalanine concentrations in the patient's home would optimise outcomes as a result of shorter turnaround times for results (Camp et al 2014; van Spronsen et al 2017). However, until the issue of test calibration and alignment between plasma and DBS specimens has been addressed, caution should be taken before adopting such technology.

Recently, Sapropterin dihydrochloride (Kuvan®), a synthetic form of the tetrahydrobiopterin cofactor for phenylalanine hydroxylase, has been used successfully to lower blood phenylalanine

concentrations. A 30% reduction is defined as being responsive to therapy (Burton et al 2007). Furthermore, lesser reductions of 10-20% may represent clinically meaningful outcomes (Burton et al 2007). However, detecting changes in phenylalanine concentrations of 10-20% using DBS samples will be challenging due to the issues outlined. It is advisable that plasma phenylalanine measurements are used to determine the response to Sapropterin therapy.

The impact of the differences in phenylalanine results obtained from different specimen types and laboratory instruments on patient management using the recommended target treatment-ranges needs to be reviewed and guidance provided in future recommendations. In addition, future guidelines should include criteria for phenylalanine test performance and ensure traceability of the method/calibration used, including the use of the DBS calibration and control materials. No study to date has derived an analytical goal by assessing the clinical impact of test performance on clinical outcomes in patients with PKU. The TAE of the test is a useful goal, as this indicates the maximum error that can be tolerated before it has a significant effect on patient management. However, the use of accurate data from published studies to derive a TAE will be limited as many state that 'blood' phenylalanine was measured and do not state whether plasma or DBS specimens were used. Furthermore, reference to the laboratory instrumentation used to quantify the 'blood' phenylalanine is rarely provided. This lack of detail of specimen type, test methodology and calibrator traceability used will impede and potentially weaken future clinical studies aimed at deriving TAE and meta-analyses assessing optimal target ranges (both upper and lower) to prevent adverse outcomes.

The results from the ERNDIM EQA pilot scheme in 2017 and 2018 indicate that there are significant problems with assay calibration resulting in an unwanted level of inter-laboratory variation and an

inherent results bias in some laboratories. This makes European/International recommendations relating to target treatment-ranges difficult to implement. The development of a commercially available CRM to standardise DBS phenylalanine tests is essential to address these issues. An international effort between professional societies, expert scientific advisory groups, PKU patient advocacy groups and organisations that have the expertise and capabilities to produce CRM material is required, in order to standardise tests.

Clinicians should take into consideration the effect of the test variability and bias (i.e. TE of the test), DBS size and quality in order to prevent over-interpretation of changes in phenylalanine concentrations, thereby preventing false re-assurances as to optimal dietary compliance. Clinical laboratories undertaking the analysis of DBS for monitoring of PKU patients should ensure that: [1] standardised criteria for the acceptance/rejection of specimens is implemented with the aim of improving the quality of DBS for monitoring; [2] FIA-MS/MS methods are replaced with LC-MS/MS methods to improve analytical performance; and [3] a rigorous evaluation of the bias between plasma and DBS phenylalanine results is undertaken in laboratories to derive a calibration factor in order to report DBS results as plasma equivalents (ideally on an individual patient basis), thereby ensuring meaningful comparison of patient results to the recommended target treatment-ranges. This is of paramount importance especially in the context of those infants that may be the subject of safeguarding measures as a result of phenylalanine concentrations that are persistently outside the target treatment-ranges, and the MU of the phenylalanine test result should be provided to guide clinicians and dieticians to allow a more objective interpretation of the monitoring of serial results. Finally, patients and parents/carers should receive regular training on blood collection techniques to ensure that more accurate and less variable results are obtained in order to achieve optimal dietary control thereby reducing adverse neurological outcomes.

References

Belton NR, Crombie JD, Robins SP, Stephen R, Farquhar JW (1973) Measurement of phenylalanine in routine care of phenylketonuric children. *Arch Dis Child* **48**:472-475.

Bonham JR & Weykamp CW (2018) ERNDIM pilot DBS EQA scheme report – Personal Communication.

Burton BK, Grange DK, Milanowski A, et al (2007) The response of patients, with phenylketonuria and elevated serum phenylalanine to treatment with oral sapropterin dihydrochloride (6R-tetrahydrobiopterin): a phase II, multicentre, open label, screening study. *J Inherit Metab Dis* **30**: 300-707.

Camp KM, Parisi MA, Acosta P, et al (2014) Phenylketonuria Scientific Review Conference: State of the science and future research needs. *Mol Genet Metab* **112**: 87-122.

Carling RS (2018a) ERNDIM Quantitative Amino Acids Method Survey, January 2018. – Personal Communication.

Carling RS (2018b) ERNDIM Quantitative Amino Acids Annual Report ERNDIM-EQAS 2017. – Personal Communication.

Chace DH, Millington DS, Terada N, et al (1993). Rapid diagnosis of phenylketonuria by quantitative analysis for phenylalanine and tyrosine in neonatal blood spots by tandem mass spectrometry. *Clin Chem* **39**: 66-71.

Chace DH, Adam BW, Smith SJ, Alexander JR, Hillman SL, Hannon WH (1999). Validation of accuracy-based amino acid reference materials in dried-blood spots by tandem mass spectrometry for newborn screening assays. *Clin Chem* **45**: 1269-1277.

CLSI. Blood Collection on Filter Paper for Newborn Screening Programs 6th ed, CLSI standard NBS01. Wayne, PA: Clinical Laboratory Standards Institute; 2013.

Corte Z & Venta R (2010) Biological variation of free plasma amino acids in healthy individuals. *Clin Chem Lab Med* **48**: 99-104.

Dhondt JL, Loeber JG, Elvers LH, Paux E (1998) Preparation of the first European working standard for phenylalanine determined in dried blood spots. *J Med Screen* **5**: 63-66.

Fonnesbeck CJ, McPheeters ML, Krishnaswami S, et al (2013) Estimating probability of IQ impairment from blood phenylalanine for phenylketonuria patients: a hierarchical meta-analysis. *J Inherit Metab Dis* **36**: 757-66.

Fraser CG (2009) Reference change values: the way forward in monitoring. *Ann Clin Biochem* **46**: 264-5.

George RS, Moat SJ (2016). Effect of dried blood spot quality on newborn screening analyte concentrations and recommendations for minimum acceptance criteria for sample analysis. *Clin Chem* **62**: 466 –75.

Gregory C, Yu C, Singh RH (2007). Blood phenylalanine for dietary compliance among patients with phenylketonuria: comparison of methods. *Genet Med* **9**: 761-765.

Groselj U, Murko S, Zerjav Tansek M et al (2015) Comparison of tandem mass spectrometry and amino acid analyzer for phenylalanine and tyrosine monitoring—implications for clinical management of patients with hyperphenylalaninemia. *Clin Biochem* **48**: 14–18.

Grunert SC, Brichta CM, Krebs A, et al (2013) Diurnal variation of phenylalanine and tyrosine concentrations in adult patients with phenylketonuria: subcutaneous microdialysis is no adequate tool for the determination of amino acid concentrations. *Nutr J* **12**: 60 doi: 10.1186/1475-2891-12-60.

Hagenfeldt L, Arvidsson A (1980) The distribution of amino acids between plasma and erythrocytes. *Clin Chem Acta* **100**: 133-41.

Hall EM, Flores SR, De Jesus VR (2015) Influence of Hematocrit and Total-Spot Volume on Performance Characteristics of Dried Blood Spots for Newborn Screening. *Int J Neonatal Screen* **1**: 69-78.

Horwitz W, Albert R (2006) The Horwitz ratio (HorRat): A useful index of method performance with respect to precision. *J AOAC Int* **89**: 1095-109.

International Organization for Standardisation (2012). Medical laboratories – requirements for quality and competence (ISO 15189:2012). Geneva: International Organization for Standardisation.

Kenny D, Fraser CG, Hyltoft Peterson P, Kallner A (1999) Strategies to set global analytical quality specifications in laboratory medicine – Consensus agreement. *Scandinavian J Clin Lab Invest* **59**: 585

Klee GG (2010) Establishment of outcome-related analytic performance goals *Clin Chem* **56**:714-722.

Lawson AJ, Bernstone L, Hall SK (2016) Newborn screening blood spot analysis in the UK: influence of spot size, punch location and haematocrit. *J Med Screen* **23**:7-16.

Lenk G, Sandkvist S, Pohanka A, Stemme G, Beck O, Roxhead N (2015) A disposable sampling device to collect volume-measured DBS directly from a fingerprick onto DBS paper. *Bioanalysis* **7**:2085-2094.

Leuthold L, Heudi O, Déglon J, Raccuglia M, Augsburger M, Picard F, Kretz O and Thomas A (2015) New Microfluidic-Based Sampling Procedure for Overcoming the Hematocrit Problem Associated with Dried Blood Spot Analysis. *Anal Chem* **87**: 2068-2071

Leuzzi V, Pansini M, Sechi E, et al (2004) Executive function impairment in early-treated PKU subjects with normal mental development. *J Inherit Metab Dis* **27**:115-25.

MacDonald A, Rylance G, Hall SK, Asplin D, Booth IW (1996). Factors affecting the variation in plasma phenylalanine in patients with phenylketonuria on diet. *Arch Dis Child* **74**: 412-417.

Mei JV, Alexander R, Adam BW, Hannon H (2001). Use of filter paper for the collection and analysis of human whole blood specimens. *J Nutr* **131**:1631-5S.

Mei JV, Zobel SD, Hall EM, De Jesus VR, Adam BW, Hannon WH (2010). Performance properties of filter paper devices for whole blood collection. *Bioanalysis* **2**: 1397-1403.

Mo XM, Li Y, Tang AG, Ren YP (2013) Simultaneous determination of phenylalanine and tyrosine in peripheral capillary blood by HPLC with ultraviolet detection. *Clin Biochem* **46**: 1074-1078.

Neto R, Gooley A, Breadmore MC, Hilder EF, Papierre F (2018) Precise, accurate and user-independent blood collection system for dried blood spot sample preparation. *Anal Bioanal Chem* **410**: 3315-3323.

Oosterhuis WP (2011) Gross overestimation of total allowable error based on biological variation. *Clin Chem* **57**: 1334-36.

Rohr FJ, Allred EN, Turner M, Simmons J, Levy HL (1996) Use of the Guthrie bacterial inhibition assay to monitor blood phenylalanine for dietary treatment of phenylketonuria. *Screening* **4**: 201-211.

Rudy JL, Rutledge JC, Lewis SL (1987) Phenylalanine and tyrosine in serum and eluates from dried blood spots as determined by reverse phase liquid chromatography. *Clin Chem* **33**:1152-1154.

Scriver CR, Kaufman S (2001). Hyperphenylalaninemia: phenylalanine hydroxylase deficiency. In: Scriver CR, Beudet AL, Sly WA, Valle D, Chils B, Kinzler KW, Vogelstein B, editors. *The Metabolic and Molecular Bases of Inherited Metabolic Disease*. 8th Ed. ed New York: McGraw-Hill, 1667-1724.

Singh RH, Rohr F, Frazier D, et al (2014) Recommendations for the nutrition management of phenylalanine hydroxylase deficiency. *Genet Med* 2014; **16**:121–131.

Smith I (1993) Recommendations on the dietary management of phenylketonuria. *Arch Dis Child* **68**:426-427.

Spooner N, Denniff P, Michielsen L, et al (2015) A device for dried blood microsampling in quantitative bioanalysis: overcoming the issues associated with blood haematocrit. *Bioanalysis* **7**:653-659.

Stroup BM, Held PK, William P, et al (2016) Clinical relevance of the discrepancy in phenylalanine concentrations analysed using tandem mass spectrometry compared with ion-exchange chromatography in phenylketonuria. *Mol Genet & Metab Reports* **6**:21-26.

van Spronsen FJ, van Wegberg AM, Ahring K, et al (2017) Key European guidelines for the diagnosis and management of patients with phenylketonuria. *Lancet Diabetes & Endocrinol* **5**:743-756.

Vockley J, Andersson HC, Antshel KM, et al (2014) Phenylalanine hydroxylase deficiency: diagnosis and management guidelines. *Genet Med* **16**: 188-200.

Wappner R, Cho S, Kronmel RA, Schuett V, Seashore MR (1999). Management of phenylketonuria for optimal outcome: A review of guidelines for phenylketonuria management and a report of surveys of parents, patients and clinic directors. *Paediatrics* **104**: e68.

Wayne PA (2012) Clinical and Laboratory Standards Institute. EP 29-A: expression of measurement uncertainty in laboratory medicine; approved guideline: Clinical and Laboratory Standards Institute.

Wendel U, Koppelkamm M, Hummel W (1991) Enzymatic phenylalanine estimation for the management of patients with phenylketonuria. *Clin Chim Acta* **201**:95-98.

Williams JP, Palmer M, Boschmans J, Lemiere F and Sobott F (2015). Gas-Phase Separation of Benzocaine Protonation Site Isomers and Phenylalanine Using Travelling-Wave Ion Mobility MS. Waters Corporation, Wilmslow, UK.

Table 1. Studies assessing the differences between plasma, DBS and whole blood phenylalanine concentrations using a variety of laboratory technologies.

Plasma assay	DBS assay	Bias DBS vs Plasma	Source
IEC	FIA-MS/MS	-19%	Gregory et al 2007
IEC	FIA-MS/MS	-28%	Grunert et al 2013
IEC	FIA-MS/MS	-26%	Groselj et al 2015
IEC	FIA-MS/MS	-28%	Stroup et al 2016
IEC	FIA-MS/MS	-18%	Present study
IEC	HPLC	-19%	Gregory et al 2007
IEC	IEC	-15%	Stroup et al 2016
Plasma assay	Whole blood assay	Bias - whole blood vs Plasma	
IEC	IEC	-7.5%	Hagenfeldt & Arvidsson 1980
HPLC	HPLC	- 8%	Mo et al 2013

Table 2 Analytical performance of the plasma and DBS tests in our laboratories against the various EQA schemes.

Sample Type	Test imprecision (CV_A) %	Test Bias %	TE of test %
Laboratory 1			
Plasma	3.0	1.6 ^a	7.5
Plasma	3.0	3.7 ^b	9.6
DBS	5.3	8 ^c	18.4
DBS	5.3	11.1 ^d	21.5
DBS	5.3	16 ^e	26.4
Laboratory 2			
Plasma	2.0	4.6 ^a	8.5
Plasma	2.0	2.1 ^b	6.0
DBS	5.1	2.9 ^c	12.9
DBS	5.1	18.2 ^d	28.2
DBS	5.1	2.7 ^e	12.7

Test bias was calculated using the all laboratory mean for the participants using the following EQA schemes; ^aUK NEQAS quantitative amino acids, ^bERNDIM plasma amino acids, ^cUK NEQAS DBS newborn screening, ^dERNDIM – DBS monitoring (Pilot scheme), ^eCDC – DBS Newborn Screening. DBS phenylalanine measured by FIA-MS/MS in both laboratories, plasma phenylalanine measured by IEC in both laboratories. TE is the total error of the test and is calculated as TE = Test bias + (1.96 x test imprecision (CV_A)).

Table 3 Inter-laboratory variation (expressed as CV%) with and without calibration using a DBS standard in 79 laboratories

Parameter	Sample 1	Sample 2	Sample 3	Sample 4	Mean
Phenylalanine concentration (μmol/L)	120	330	420	940	452
Inter-laboratory CV without calibration	23%	20%	18%	20%	20%
Inter-laboratory CV with calibration	14%	9%	8%	9%	10%